US009323892B1

US 9,323,892 B1

(12) **United States Patent**
Paris, III

(10) **Patent No.:** **US 9,323,892 B1**
(45) **Date of Patent:** *Apr. 26, 2016

(54) **USING DE-IDENTIFIED HEALTHCARE DATA TO EVALUATE POST-HEALTHCARE FACILITY ENCOUNTER TREATMENT OUTCOMES**

(71) Applicant: **Vigilytics LLC**, Victor, NY (US)

(72) Inventor: **Andrew L. Paris, III**, Victor, NY (US)

(73) Assignee: **Vigilytics LLC**, Victor, NY (US)

(*) Notice: Subject to any disclaimer, the term of this patent is extended or adjusted under 35 U.S.C. 154(b) by 0 days.

This patent is subject to a terminal disclaimer.

(21) Appl. No.: **14/082,433**

(22) Filed: **Nov. 18, 2013**

Related U.S. Application Data

(63) Continuation-in-part of application No. 12/827,745, filed on Jun. 30, 2010.

(60) Provisional application No. 61/222,428, filed on Jul. 1, 2009, provisional application No. 61/729,207, filed on Nov. 21, 2012.

(51) **Int. Cl.**
G06F 21/62	(2013.01)
G06F 19/00	(2011.01)
G06Q 50/22	(2012.01)

(52) **U.S. Cl.**
CPC *G06F 19/322* (2013.01); *G06F 19/366* (2013.01); *G06F 21/6245* (2013.01); *G06Q 50/22* (2013.01)

(58) **Field of Classification Search**
CPC . G06F 19/322; G06F 19/366; G06F 21/6245; G06Q 50/22

USPC 726/26, 7, 9, 17, 19, 2.7
See application file for complete search history.

(56) **References Cited**

U.S. PATENT DOCUMENTS

2004/0143403 A1*	7/2004	Brandon et al.	702/19
2005/0234739 A1	10/2005	Schoenberg	
2005/0261941 A1	11/2005	Scarlat	
2006/0004772 A1*	1/2006	Hagan et al.	707/10
2006/0020611 A1*	1/2006	Gilbert et al.	707/100
2006/0184524 A1	8/2006	Pollanz	
2006/0229919 A1	10/2006	Pugh	
2007/0043594 A1*	2/2007	Lavergne	705/2
2007/0219824 A1*	9/2007	Rawlings et al.	705/2
2008/0010254 A1*	1/2008	Settimi	707/3
2008/0040151 A1*	2/2008	Moore	705/2

(Continued)

OTHER PUBLICATIONS

"Health Information Privacy: Business Associates," U.S. Department of Health and Human Services, accessed online http://www.hhs.gov/ocr/privacy/hipaa/understanding/coveredentities/businessassociates.html, 2 pages.
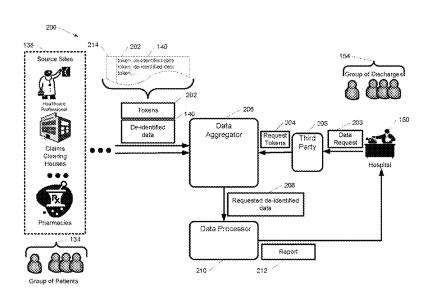
(Continued)

Primary Examiner — Catherine Thiaw
Assistant Examiner — Don Zhao
(74) *Attorney, Agent, or Firm* — Fish & Richardson P.C.

(57) **ABSTRACT**

A computer-implemented method includes producing medical information that characterizes a group of individuals from a set of private data representing pre or post-encounter characteristics of the individuals, wherein the individuals have had encounters with a healthcare facility. The identity of the individuals is unattainable from the produced medical information. The method also includes providing the produced medical information to report the pre or post-encounter characteristics of the group.

26 Claims, 7 Drawing Sheets

(56)   **References Cited**

U.S. PATENT DOCUMENTS

2008/0120296 A1*   5/2008 Kariathungal et al. 707/6
2008/0133273 A1*   6/2008 Marshall 705/3
2008/0147554 A1*   6/2008 Stevens et al. 705/51
2008/0281631 A1* 11/2008 Syth et al. 705/2
2009/0024416 A1   1/2009 McLaughlin et al.
2009/0070146 A1   3/2009 Haider et al.
2009/0287502 A1* 11/2009 Roberts et al. 705/3
2010/0162355 A1*   6/2010 Zimmerman et al. 726/2
2011/0184748 A1*   7/2011 Fierro et al. 705/2
2011/0225007 A1   9/2011 Theis
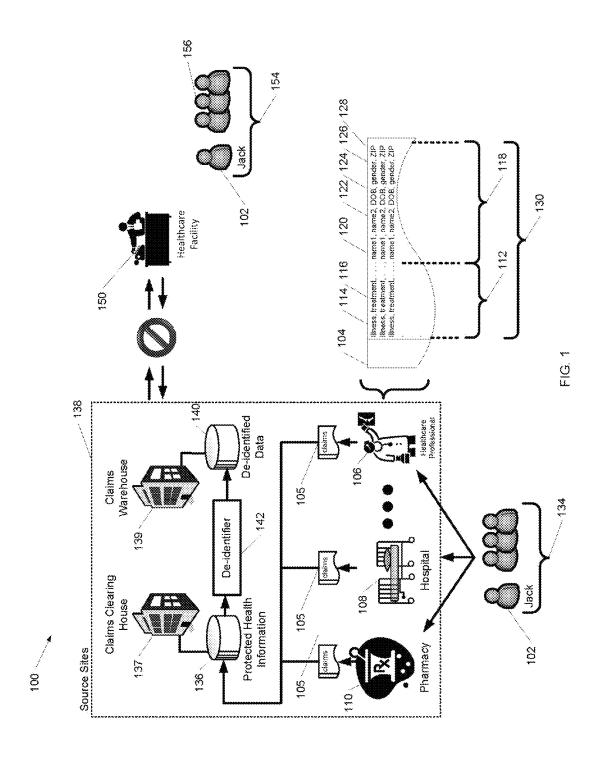
2011/0258000 A1   10/2011 Green et al.
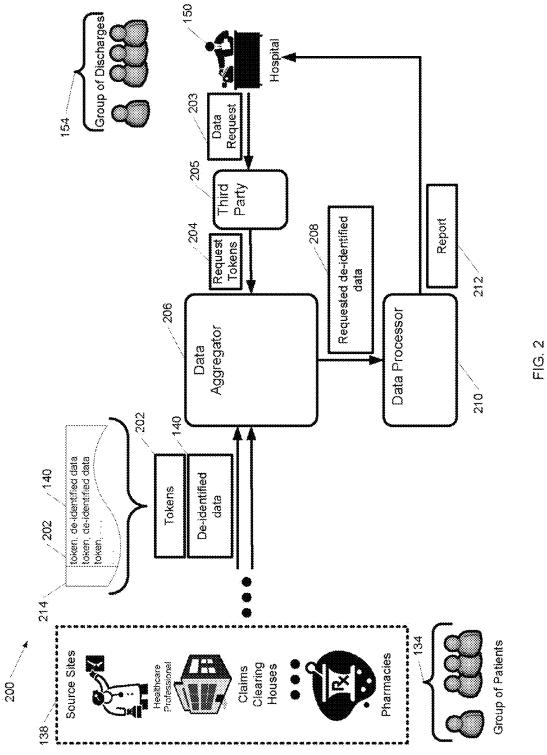
OTHER PUBLICATIONS

"Group," The American Heritage Dictionary, 2014 Houghton Mifflin Harcourt, Retrieved online: https://ahdictionary.com/word/search. html?q=group (accessed Oct. 23, 2014), 3 pages.
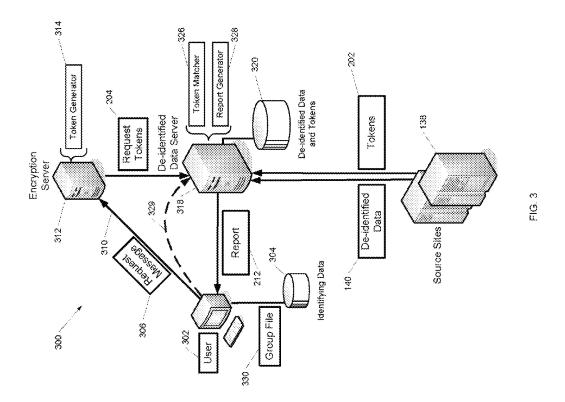"Health Information Privacy: Business Associates," U.S. Department of Health and Human Services, Apr. 3, 2003, Retrieved online: http://www.hhs.gov/ocr/privacy/hipaa/understanding/coveredentities/businessassociates.html, (accessed Oct. 9, 2014), 2 pages.

* cited by examiner

FIG. 1

FIG. 2
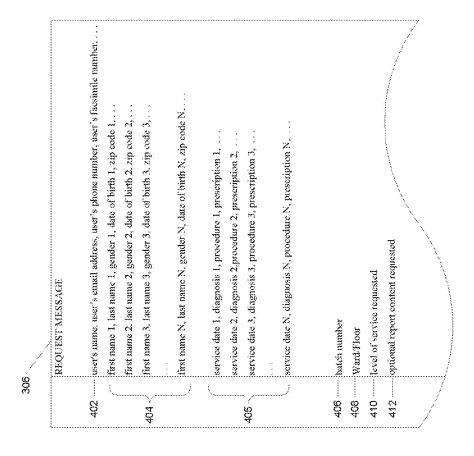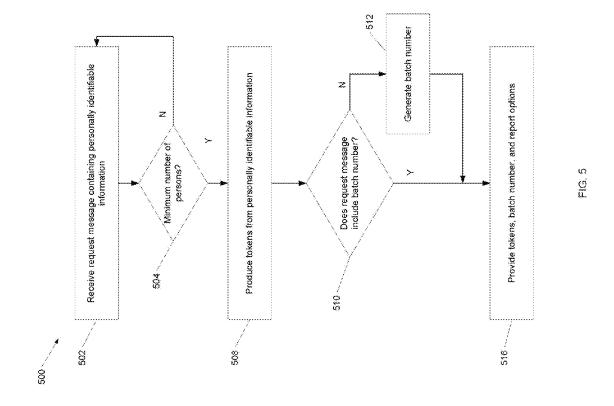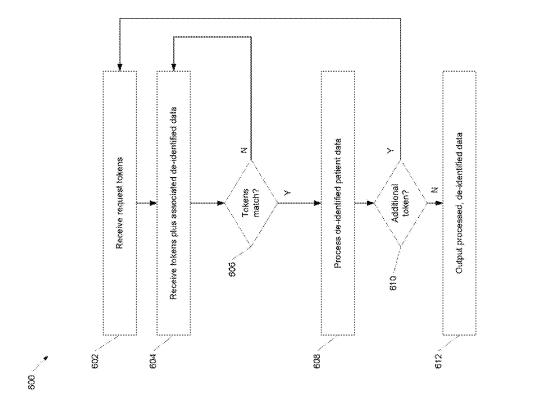
FIG. 3

306

REQUEST MESSAGE

402 —— user's name, user's email address, user's phone number, user's facsimile number. . . . .

404 {
first name 1, last name 1, gender 1, date of birth 1, zip code 1. . . .
first name 2, last name 2, gender 2, date of birth 2, zip code 2, . . .
first name 3, last name 3, gender 3, date of birth 3, zip code 3, . . .
:
first name N, last name N, gender N, date of birth N, zip code N . . . .
}

405 {
service date 1, diagnosis 1, procedure 1, prescription 1, ...
service date 2, diagnosis 2, procedure 2, prescription 2, ...
service date 3, diagnosis 3, procedure 3, prescription 3, ...
:
service date N, diagnosis N, procedure N, prescription N, ...
}

406 —— batch number

408 —— Ward/Floor

410 —— level of service requested

412 —— optional report content requested

FIG. 4

FIG. 5

600

602 — Receive request tokens

604 — Receive tokens plus associated de-identified data

606 — Tokens match?
 N
 Y

608 — Process de-identified patient data

610 — Additional token?
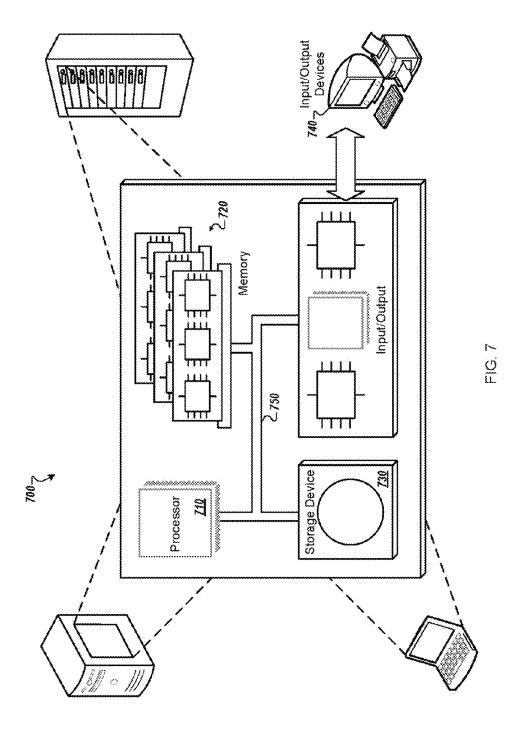 Y
 N

612 — Output processed, de-identified data

FIG. 6

FIG. 7

USING DE-IDENTIFIED HEALTHCARE DATA TO EVALUATE POST-HEALTHCARE FACILITY ENCOUNTER TREATMENT OUTCOMES

CLAIM OF PRIORITY

This is a continuation-in-part application of U.S. application Ser. No. 12/827,745, filed Jun. 30, 2010 which claims priority to U.S. provisional application No. 61/222,428, filed on Jul. 1, 2009; this application also claims priority under 35 USC §119(e) to U.S. Patent Application Ser. No. 61/729,207, filed on Nov. 21, 2012; the entire contents of each are hereby incorporated by reference.

BACKGROUND

The present disclosure relates to processing and transmitting personal data, the dissemination of which is restricted by federal law.

Due to federal privacy laws and the inconsistency of electronic medical record deployments, healthcare facilities (e.g., hospitals, nursing homes, residential mental health facilities, prison medical wards) are often unable to obtain information about health care services patients receive before or after healthcare facility encounters. Having access to these data may allow healthcare facilities to better serve patients. Gaining an understanding of care that occurs in the outpatient setting is needed to evaluate potential for changes in the process of inpatient care. Not having this information may lead to unrecognized suboptimal care and ultimately to healthcare facility readmission for which the healthcare facility may be held accountable. Outpatient healthcare providers such as pharmacies and physicians generate private healthcare data about patients, including medical and prescription drug data, and administrative healthcare claims data. Data that associates patient identity with health information is known as protected health information (PHI). Healthcare providers can store protected health information in electronic databases for future use in patient care and insurance claims processing. Healthcare facilities could identify changes in the process of care provided (e.g., patient and provider education, adherence follow-up) by having information on trends in health care utilization before and after health care facility encounters. Federal privacy laws, however, make obtaining this information a cumbersome process which would add considerable expense to operations.

SUMMARY

The systems and techniques described here relate to de-identifying patient identities from healthcare facility encounters and retrieving information from pre or post encounter activity.

In one aspect, a computer-implemented method includes producing medical information from a set of private data representing pre or post-encounter characteristics of the individuals, where the individuals have had an encounter with a healthcare facility. The identity of the individuals is unattainable from the produced medical information. The method includes using the produced medical information to report the pre or post-encounter characteristics for the group.

Implementations may include any of all of the following features. Producing medical information that characterizes the group of encounters may include producing a request token for each individual included in the group of encounters. Producing such a request token for each individual may

include encrypting respective data that identifies each individual. Producing medical information that characterizes the group of encounters may include comparing the request tokens to tokens associated with the medical information to be produced. The tokens associated with the medical information to be produced and the request tokens may be similarly encrypted. Producing medical information that characterizes the group of encounters may include determining if the comparison provides at least a minimum number of matches. Producing medical information that characterizes the group may include requesting a predefined portion of the medical information. Additionally, producing medical information that characterizes the group may include determining if the group includes at least a minimum number of individuals. The private data may represent medical information associated with pre or post-encounter medical activity for the individuals within the group. The private data may be obtained from pharmacies or other medical care providers.

In another aspect a system includes an encryption server for producing a request token for each individual included in a group identified in a request for medical information that characterizes the group consisting of individuals who had an encounter with a healthcare facility. The system may also include a data server for producing the medical information that characterizes the group from a set of private data representing pre or post-encounter characteristics of the individuals. The identity of the individuals is unattainable from the produced medical information. The data server is also configured to provide the produced medical information to report the characteristics of the group of encounters.

Implementations may include any of all of the following features. The data server may provide a request token for each individual included in the group of encounters to produce the medical information that characterizes the group of encounters. The request token for each individual may represent encrypted data that identifies the corresponding individual. The data server may be configured to compare the request tokens to tokens associated with the medical information to be produced. The tokens associated with the medical information to be produced and the request tokens may be similarly encrypted. The data server may be configured to determine if the comparison provides at least a minimum number of matches. The request may represent a predefined portion of information to use for producing the medical information that characterizes the group of encounters. The encryption server may be configured to determine if the group includes at least a minimum number of individuals. The private data may represent medical information associated with pre or post-encounter medical activity of the individuals within the group. The private data may be obtained from pharmacies or other medical care providers.

In another aspect, one or more computer readable media storing instructions that are executable by a processing device, and upon such execution cause the processing device to perform operations that include producing medical information that characterizes a group of individuals from a set of private data representing pre or post encounter characteristics of the individuals. The individuals have had an encounter with a healthcare facility. The identity of the individuals is unattainable from the produced medical information. The operations also include providing the produced medical information to report the characteristics of the group.

Implementations may include any of all of the following features. Producing medical information that characterizes the group may include producing a request token for each individual included in the group. Producing such a request token for each individual may include encrypting respective

data that identifies each individual. Producing medical information that characterizes the group of encounters may include comparing the request tokens to tokens associated with the medical information to be produced. The tokens associated with the medical information to be produced and the request tokens may be similarly encrypted. Producing medical information that characterizes the group may include determining if the comparison provides at least a minimum number of matches. Producing medical information that characterizes the group may include requesting a predefined portion of the information. Additionally, producing medical information that characterizes the group of encounters may include determining if the group includes at least a minimum number of individuals. The private data may represent medical information associated with pre or post-encounter medical activity of the individuals of the group of encounters. The private data may be obtained from pharmacies or other medical care providers.

The details of one or more embodiments of the invention are set forth in the accompanying drawings and the description below. Other features, objects, and advantages of the invention will be apparent from the description and drawings, and from the claims.

DESCRIPTION OF DRAWINGS

FIG. 1 illustrates exemplary circumstances in which protected health information is stored by healthcare providers, claims clearing houses, and other source sites, and requested by a hospital.

FIG. 2 illustrates an exemplary method and system that enables a hospital to obtain medical information about a group of people without violating the privacy of persons in the group.

FIG. 3 illustrates an exemplary system incorporating an encryption server and a de-identified data server to enable a user who is not permitted to obtain the private data of a group of individuals who have been discharged from a hospital to instead obtain a report that characterizes the group of discharges as a whole.

FIG. 4 illustrates an exemplary message requesting de-identified data about a group of discharges.

FIG. 5 is a flowchart that represents exemplary operations of a token generator.

FIG. 6 is a flowchart that represents exemplary operations of a token matcher.

FIG. 7 represents a computer system and related components.

Like reference symbols in the various drawings indicate like elements.

DETAILED DESCRIPTION

Referring to FIG. 1, when a doctor treats Jack 102 for high blood pressure, a surgeon removes Jack's gallstones, and a pharmacist fills Jack's prescription for insulin, Jack leaves a trail of electronic records 104 with healthcare professionals (e.g., healthcare professional 106), in their offices (e.g., doctor's office, nurse's station, etc.), in healthcare facilities (e.g., a hospital 108, a pharmacy 110, a nursing home, etc.) and the like. The electronic records 104 contain medical data 112 about Jack, for example his illnesses 114 and treatments 116. Each piece of medical data 112 is associated with personally identifiable information 118 that identifies Jack and distinguishes him from all other patients, such as his first name 120, last name 122, date of birth 124, gender 126, and zip code

128. Together, the pieces of medical data 112 and the personally identifiable information 118 make up Jack's protected health information (PHI) 130.

Each of Jack's healthcare providers may submit health insurance claims 105 containing PHI 130, as well as additional PHI 130, to a claims clearing house 137. The claims clearing house 137 may store the PHI 130 of many patients 134, including Jack, in a PHI database 136. A de-identifier 142 can process the PHI 130 to generate irreversibly de-identified data 140 by removing all personally identifiable information 118 or otherwise transforming the PHI 130 so that it cannot be associated with a particular person. A claims warehouse 139 stores de-identified data 140 about many patients 134. Claims clearing houses 137 and claims warehouses 139 are optimized for retrieving and providing PHI 136 and de-identified data 140 for use in further processing, but health care providers such as healthcare professionals 106 and pharmacies 110 may also serve as source sites 138 for de-identified data in a distributed system.

By searching for Jack's personally identifiable information 118 in its electronic records 104, the pharmacy 110 is able to look up the various drugs in Jack's PHI 130 and, for example, check for bad interactions among them. Jack does not object to this use of his PHI 130 because it improves the quality of his healthcare. On the other hand, Jack would object to the pharmacy 110 giving his PHI 130 to third parties without his permission because that would disclose personal, private information about him that Jack's potential employers, for example, might use to discriminate against him.

Federal laws, such as the Health Insurance Portability and Accountability Act (HIPAA), protect Jack by prohibiting the source sites 138 possessing Jack's PHI 130 from releasing it to third parties without Jack's permission. Under HIPAA, the source sites 138 can release only irreversibly de-identified data 140 without Jack's permission.

A healthcare facility, for example a hospital 150, may be interested in outpatient medical information contained in Jack's PHI 130 that occurs post-discharge. Since the hospital may not have access to post discharge medical information it may not have the knowledge to fully understand the potential risk of rehospitalizations or other medical events for a group 154. In that case, the hospital 150 would like to use the PHI 136 from the group 154 to assess the risk of rehospitalization or other medical events. The hospital can use this information to make adjustments to treatments and improve the quality of care. Unfortunately, the group 154 may also be too large for the hospital 150 to practically obtain permission from each person in the group, without which HIPAA prohibits the healthcare providers 138 from releasing the PHI 136 to the hospital 150 for care not related to the hospitalization. Practically speaking, even when hospitals have the authority to obtain the requested data they may not have a mechanism for doing so. The de-identified data 140, which the healthcare providers or other source sites 138 could release without permission, is not useful to the hospital 150 because the hospital has no way to know whether it corresponds to the people in the group 154. With no way to obtain pre and post hospitalization medical data for the group, the hospital 150 cannot determine post discharge medical events for the group.

Referring to FIG. 2, to provide the hospital 150 with the pre and post hospitalization medical data necessary to generate the post discharge report for the group 154, without revealing the protected health information of persons in the group 154, the exemplary system 200 associates unique tokens 202 with the irreversibly de-identified data 140. The tokens 202 correspond to persons treated by healthcare providers but do not reveal the identities of the treated individuals. The hospital

150 can send a data request 203 to a third party 205, who in turn can generate request tokens 204, one for each person in the group 154. The third party 205 can send the request tokens 204 to a data aggregator 206. The data aggregator 206 stores de-identified data 140 and associated tokens 202. By searching for the request tokens 204 among the stored tokens 202, the data aggregator 206 can process the de-identified data to generate requested de-identified data 208 that corresponds to the group 154. A data processor 210 can process the requested de-identified data 208 to generate a report 212 containing metrics such as medication utilization reports that are useful to the hospital 150 in assessing the overall post discharge health status of the group 154.

By using the tokens 202 and request tokens 204, no parties other than the authorized source sites 138 can associate de-identified data 140 with the identity of any person in the group of patients 134 or discharges (e.g., the group 154). The hospital 150 requesting the report 212 can never receive data associated with individuals. Further, the information in the report 212 may be processed into metrics that characterize a large group and cannot be used to infer information about individuals. De-identified data records may be provided to the data processor 210 (e.g., a third party data processor), but that third party may not have access to any personally identifiable information 118 about the group 154. Nor may any party with access to the de-identified data 140, other than the healthcare providers 138, also have access to the de-identifier 142. These features of the system 200 maintain the privacy of the protected health information 130.

While we describe a system in which a hospital 150 needs to estimate the post discharge health status of a group of discharges, the system 200 can work for applications in which information characterizing a group needs to be generated from the private data of group members. In one arrangement, the system may implement Microsoft Windows-based computers in connection with internet-based components. However, other implementations may use other types of components that support the processing of pre and post hospitalization medical data from healthcare databases.

While the hospital 150 is referred to as a location interested in the post discharge health status of a group of discharges, one or more other types of healthcare facilities may utilize the techniques and methodologies described herein. Such healthcare facilities may include dialysis centers, specialty clinics, retirement homes, rehabilitation centers, or other types of entities where post encounter follow up may be of interest.

Referring to FIG. 3, in an exemplary system 300, a user 302 specifies a group of people using identifying data for each group member. Referring briefly to FIG. 4, an exemplary set of identifying data 404 is illustrated. Returning to FIG. 3, the user may obtain the personally identifiable information from a database 304 by retrieving it in the form of a group file 330. The user 302 generates a request message 306, which may be a computer file, which uniquely identifies each member of the group using the identifying data 304. Typically, the group contains at least a minimum number of unique members. A request message 306 may be generated at a remote computer operated by the user 302.

The request message 306 contains a set of identifying data 404 (see FIG. 4) for each group member, as well as data that identifies a user 402 (see FIG. 4), such as the user's name, email address, phone number and facsimile number, and a batch number 406 that identifies the request. The group members' identifying data 404 (see FIG. 4) may include first name, last name, date of birth, gender, and zip code as well as other identifiers, such as social security number, that may be added to ensure that members are uniquely identified. In some

examples, fewer identifiers may be used. The request message 306 may also include optional indications that the user 302 requests a particular type of report 212 or requests additional processing to enhance the value of the report. The request message may also contain additional non-identified (not containing any personally identifiable information) data 405 to be used by a report generator 328. These additional non-identified data 405 will be encrypted upon submission allowing access to only a report generator 328. Alternatively, the additional data 405 may be passed directly to the report generator 328 as shown by dotted line 329. These non-identified data 405 may include service dates, diagnosis information, or other pre-encounter and/or post-encounter details.

The request message 306 is sent over a communications network 310 (e.g., the Internet, a LAN, etc.) to an encryption server 312. If the request message 306 contains a minimum number of unique group members, the encryption server 312 creates a unique token, or identifier, for each person in the message 306. The encryption server creates each token by applying a token generator 314 that encrypts the personally identifiable information of each group member and passes the pre-encrypted non-identified data through the system to a report writer. The set of tokens corresponding to all the group members constitutes a batch of request tokens 204. The minimum number of group members, for example ten, is chosen to make it effectively impossible to associate individual group members with individual tokens in the batch of request tokens 204.

The encryption server 312 provides the request tokens 204 to the de-identified data server 318. The de-identified data server 318 stores records of de-identified data and corresponding tokens 320 obtained from source sites 138 such as pharmacies, healthcare professionals, rehabilitation centers, and electronic claims clearing houses. Other sources of data can include prison records, arrest records, the national death index or other death records, medical laboratories, dialysis centers, nursing homes, radiology providers, potentially military records, employment records, financial records, or DMV records. Each token obtained from the source sites 138 may have been created using the same token generator 314 used by the encryption server 312, or using any other means that generates the identical token for the same personally identifiable information.

For example, a pharmacy tracks the prescription histories of the patients being served. De-identified prescription histories and corresponding tokens are sent to the de-identified data server 318. At the request of a user 302, such as a hospital, the encryption server 312 generates tokens identical to the patients' tokens using the same personally identifiable information. A unique token corresponding to the same personally identifiable information permits the pharmacy and the hospital to refer to the same anonymous people without the hospital ever associating protected health information with a particular person. To produce tokens and request tokens, one or more encryption techniques may be utilized, for example, hash functions and other methodologies may be implemented.

A token matcher 326, executed by the de-identified data server 318, performs a look-up in the de-identified database 320 to find all tokens in the database that match the tokens sent from the encryption server 312. All available data for matched tokens, the requested de-identified data 208, is retrieved for use in the report generator 328. The de-identified data may include, for example, prescription data, medical claims, and hospital claims. The de-identified data server 318 or the data processor 210 may use the report generator 328 to process the requested de-identified data 208 in a way that

leaves it irreversibly de-identified. An example of such processing is an algorithm that transforms the data into a post discharge health status assessment. The report generator formats the processed data into an electronic or hardcopy report **212** that is returned to the user.

While the report **212** is described as generated on a computer system, it may also be generated in part or entirely outside the computer system. For example, the report **212** could be conveyed to the user via regular mail or other similar technique. In particular, the report may be generated and printed at the site of the de-identified data server **318** and subsequently communicated to the user **302** without using the computer network. The report generator **328** may also reside in the data processor **210** separate from the de-identified data server **318**.

In one arrangement, once the relevant information and options have been selected, the user **302** submits the request by clicking a submit button. The request message **306** may be encrypted prior to being transmitted over a computer network **310**. At the encryption server **312**, the request message **306** is unencrypted and stored. The encryption server **312** may send an optional confirmation message to the user **302**. The confirmation message may include the time and date that the message was received, and may indicate the service level and options selected by the user **302**.

Referring to FIG. **4**, the user **302** can input the user's identifying information **402**, including name, email address, phone number and/or facsimile number, in the first record, or set of fields, of the request message **306**. The user can then input the personally identifiable information **404** for each of the group members, including first name, last name, gender, date of birth, and zip code. The user may also input additional data **405** such as service dates, diagnosis information, or other encounter details that will be rendered non-identified and encrypted upon submission. Equivalently, the user can incorporate the group file **330** into the request message **306**. Information about the ward **408** from which the group was discharged, or other purpose for which the report **212** is being requested may also be included in the request message **306**. The format of the request message **306** may be adjusted based on the user's **302** needs.

The request message **306** may also include information about the level of service **410** requested by the user **302**, for example the quantity, quality or type of information. A first level of service may request up to six months of pre and post hospitalization medical data; a second level may request up to twelve months of data; and a third level may request a two year pre and post hospitalization medical data. Alternatively, instead of providing the user **302** with a variety of service level options **410**, the system **300** may simply retrieve all of the pre and post hospitalization medical data available for the group.

The user **302** may also request additional, optional information **412**. For example, the user **302** may request information regarding the drug categories and drug indications associated with the drugs in the de-identified data. Drug indications include the medical conditions associated with each drug. Drug categories include the type of drug. This data can be passed to the data processor **210** to include in the report **212**. Alternatively, this data may be returned as part of every report **212**.

Referring to FIG. **5**, a flowchart **500** represents a particular arrangement of operations of a token generator (e.g., the token generator **314** shown in FIG. **3**). Operations include receiving **502** the request message (e.g., request message **306**) containing the personally identifiable information of each person in the group (e.g., such as the group **154** shown in

FIG. **1**). Upon receiving the request message, operations may also include determining **504** whether the message includes the minimum number of persons. If the message does not include the minimum number of persons, the token generator returns to receive **502** a request message. A limit on the minimum number of persons ensures that a report (e.g., the report **212**) covers enough people such that it is difficult (if not impossible) to infer any association between particular persons and the information in the report. If the message does include the minimum number of people, operations of the token generator include producing **508** tokens from the identifying data of each person. Producing the tokens may include applying an encryption algorithm to the identifying data of each person. The tokens uniquely identify each person in the group **154**.

Operations also include determining **510** if the request message includes a batch number. If the message does not include a batch number, operations include generating **512** a batch number. Operations also include providing **516** the tokens, batch number and report options. The individual tokens are placed in a batch file and may be encrypted before being transmitted over the network.

Each batch file of request tokens **204** also specifies the information needed for the report **212**. The request tokens are transmitted to the de-identified data server **318**, unencrypted and processed using rules for searching, matching, and retrieving healthcare data.

Referring to FIG. **6**, a flowchart **600** represents a particular arrangement of operations of a token matcher (e.g., the token matcher **326** shown in FIG. **3**). Operations include receiving **602** request tokens (e.g., such as the request tokens **204** shown in FIG. **2**). Operations may also include receiving **604** a token (e.g., such as the token **202**) and associated de-identified data (e.g., such as the de-identified data **140**), typically from source sites (e.g., such as the source site **138** shown in FIG. **1**). Upon receiving a request token, de-identified data and a token, operations include determining **606** whether the request token matches the token. If no match is found, operations include returning to receive **604** more tokens and de-identified data. If a match is found, operations include processing **608** the de-identified data. Processing may include storage for later retrieval. In some arrangements, operations may also include determining whether a minimum number of matches has been detected. If such a minimum number of matches has not occurred in comparing the request tokens and the tokens, action may be taken (e.g., pause or restart processing) until the predefined number of matches has been detected (e.g., so as not to increase the probability of one or more individuals being identified by a process of elimination).

In addition to determining that a minimum number of matches has been found operations may also include processing of other privacy-preserving rules. For example, a rule may not allow the same user **302** (e.g., hospital **150**) to submit more than one request message **306** containing substantially the same identifying data **404** within a defined period of time, e.g., six months. Identifying substantially the same identifying data **404** from a user **302** may involve comparing the identifying data **404** from different stored request messages **306** and checking that at least a minimum variation exists in the request messages **306**. For example, this may involve checking that the identifying data **404** for at least five patients does not match the identifying data **404** for any previous search request **306** by the user **302**. Especially when combined with the size requirements and less than 100% capture rate, such rules greatly reduce the possibility of identifying any one person's data.

The de-identified data (e.g., such as the de-identified data **140**) may include a list of drugs prescribed over the requested period for the members of the group (e.g., such as the group **154**). The list of drugs prescribed may include the drug name, form, strength, days supplied, and date dispensed. The de-identified data could also include data from other sources such as those listed above, and include diagnosis, procedures, prison release date, date of death, date of arrest, etc. As part of processing **608** the de-identified data, operations may include determining the drug category and drug indications for each drug prescribed. Operations may also include accessing a database relating the drug category and indications to each possible drug. The database may be maintained within the de-identified data server **318** database, or may be accessed on a remote server maintained by a third party.

Upon processing the de-identified data, operations also include determining **610** if additional tokens remain in the batch of request tokens (e.g., such as the request tokens **204**). If additional tokens remain, operations include receiving **602** more request tokens. If there are no additional request tokens, operations may include outputting **612** the processed, de-identified data.

Operations may also include providing the number of tokens submitted, the number matched, and the overall match rate. The collected data covers the interval of historical data according to the level of service requested (e.g., as represented by the level of service **410** in FIG. **4**). De-identified diagnosis and procedures data may also be included from administrative medical claims from a doctor's office (e.g., office **106**) or a healthcare facility (e.g., the hospital **108**).

In addition to accessing and incorporating drug indication information for each drug prescribed to persons in the group (e.g., such as group **154**), operations may include further processing of the requested de-identified data. For instance, operations may include determining the probability that a particular drug indicates a particular condition. In this example, in addition to providing the possible indications, the requested de-identified data would include the likelihood that anonymous individuals associated with the request tokens (e.g., such as request tokens **204**) have each of the conditions indicated by the prescribed drugs. Operations may also include using expert rule systems to provide health status information based on the prescription drug medical data. Alternatively, operations may include using diagnosis codes from medical claims data to assess health status. In some arrangements, operations may also include integrating the de-identified data with the non-identified data **405**.

The requested de-identified data **208** may be sent for further processing to a third party data processor (e.g., data processor **210**) who may apply proprietary algorithms, modify the data format, or generate additional reports, provided that no re-identifiable information is transmitted to the user **302**. Third parties may not have access to the request message **306** and the group file **330** so that no association may be inferred between the de-identified data **140** and particular persons in the group **154**.

In some implementations, the report **212** can include analysis results. The analysis results can include statistical analyses, correlations of varying post-discharge characteristics, comorbidity rates, or other metadata. The analysis results can be determined by the third party data processor. The analysis results can be generated automatically, or by human intervention. For example, in addition to the report **212**, a human analyst can respond to the user **302** with custom analysis relating to the user's **302** request message **306**.

The report **212** provides the hospital **150** with information for making an informed decision about the relationship

between the hospitalization and the health status of the group of discharges. Hospitals could identify changes in the process of care provided during or after the hospitalization by having information on trends in outpatient health care utilization. In particular, the hospital **150** may change internal policies including discharge instructions, structure of care, process of care, or other relevant operating procedures based on the information in the report **212**. Patterns from multiple reports over time may be used to determine which actions to take, if any. The report **212** may be used alone to make decisions about the health status of the group **154**, or may simply indicate that additional investigation is needed.

In some implementations, all reports **212** can be saved and archived by the system **200**. Data on the saved reports **212** can be made available to the hospital **150**. For example, a hospital **150** may query the system **200** about a group **154** of patients sharing a characteristic, and deliver a report for the group as well as a second report. The second group could include, for example, comparative data for all patients discharged from the hospital.

FIG. **7** is a schematic diagram of a generic computer system **700**. The system **700** can be used for the operations described in association with any of the computer-implemented methods described previously, according to one implementation. The system **700** includes a processor **710**, a memory **720**, a storage device **730**, and an input/output device **740**. Each of the components **710**, **720**, **730**, and **740** are interconnected using a system bus **750**. The processor **710** is capable of processing instructions for execution within the system **700**. In one implementation, the processor **710** is a single-threaded processor. In another implementation, the processor **710** is a multi-threaded processor. The processor **710** is capable of processing instructions stored in the memory **720** or on the storage device **730** to display graphical information for a user interface on the input/output device **740**.

The memory **720** stores information within the system **700**. In some implementations, the memory **720** is a computer-readable medium. The memory **720** is a volatile memory unit in some implementations and is a non-volatile memory unit in other implementations.

The storage device **730** is capable of providing mass storage for the system **700**. In one implementation, the storage device **730** is a computer-readable medium. In various different implementations, the storage device **730** may be a floppy disk device, a hard disk device, an optical disk device, or a tape device.

The input/output device **740** provides input/output operations for the system **700**. In one implementation, the input/output device **740** includes a keyboard and/or pointing device. In another implementation, the input/output device **740** includes a display unit for displaying graphical user interfaces.

The features described can be implemented in digital electronic circuitry, or in computer hardware, firmware, software, or in combinations of them. The apparatus can be implemented in a computer program product tangibly embodied in an information carrier, e.g., in a machine-readable storage device, for execution by a programmable processor; and method steps can be performed by a programmable processor executing a program of instructions to perform functions of the described implementations by operating on input data and generating output. The described features can be implemented advantageously in one or more computer programs that are executable on a programmable system including at least one programmable processor coupled to receive data and instructions from, and to transmit data and instructions to,

a data storage system, at least one input device, and at least one output device. A computer program is a set of instructions that can be used, directly or indirectly, in a computer to perform a certain activity or bring about a certain result. A computer program can be written in any form of programming language, including compiled or interpreted languages, and it can be deployed in any form, including as a stand-alone program or as a module, component, subroutine, or other unit suitable for use in a computing environment.

Suitable processors for the execution of a program of instructions include, by way of example, both general and special purpose microprocessors, and the sole processor or one of multiple processors of any kind of computer. Generally, a processor will receive instructions and data from a read-only memory or a random access memory or both. The essential elements of a computer are a processor for executing instructions and one or more memories for storing instructions and data. Generally, a computer will also include, or be operatively coupled to communicate with, one or more mass storage devices for storing data files; such devices include magnetic disks, such as internal hard disks and removable disks; magneto-optical disks; and optical disks. Storage devices suitable for tangibly embodying computer program instructions and data include all forms of non-volatile memory, including by way of example semiconductor memory devices, such as EPROM, EEPROM, and flash memory devices; magnetic disks such as internal hard disks and removable disks; magneto-optical disks; and CD-ROM and DVD-ROM disks. The processor and the memory can be supplemented by, or incorporated in, ASICs (application-specific integrated circuits).

To provide for interaction with a user, the features can be implemented on a computer having a display device such as a CRT (cathode ray tube) or LCD (liquid crystal display) monitor for displaying information to the user and a keyboard and a pointing device such as a mouse or a trackball by which the user can provide input to the computer.

The features can be implemented in a computer system that includes a back-end component, such as a data server, or that includes a middleware component, such as an application server or an Internet server, or that includes a front-end component, such as a client computer having a graphical user interface or an Internet browser, or any combination of them. The components of the system can be connected by any form or medium of digital data communication such as a communication network. Examples of communication networks include, e.g., a LAN, a WAN, and the computers and networks forming the Internet.

The computer system can include clients and servers. A client and server are generally remote from each other and typically interact through a network, such as the described one. The relationship of client and server arises by virtue of computer programs running on the respective computers and having a client-server relationship to each other.

A number of implementations have been described. Nevertheless, it will be understood that various modifications may be made without departing from the spirit and scope of the following claims.

What is claimed is:

1. A computer-implemented method comprising:

receiving a request from a user device for medical information that characterizes a group of individuals composed of multiple individuals from a set of de-identified medical data representing pre- or post-encounter characteristics of the individuals;

producing, by a hardware processor, tokens, wherein the identity of the individuals is unattainable from the

tokens, each token being associated with an individual in the group, each token associable with a corresponding token in the set of de-identified medical data, and each token in the set of de-identified medical data being associated with medical data for a corresponding individual;

producing the requested medical information that characterizes the group of individuals by identifying the tokens in the set of de-identified medical data and the associated de-identified medical data that correspond to the produced tokens, wherein the produced medical information is absent personally identifiable information and has been produced without the individuals in the group having authorized production of such medical information that characterizes the group; and

providing the produced medical information to report the characteristics of the group of individuals absent authorization from the individuals in the group of individuals.

2. The computer-implemented method of claim 1, in which producing the token for each individual includes encrypting respective data that identifies each individual.

3. The computer-implemented method of claim 1, in which producing the requested medical information that characterizes the group includes comparing the produced token for each individual in the group to the tokens associated with the set of medical data.

4. The computer-implemented method of claim 3, in which the tokens associated with the set of medical information and the produced token for each individual in the group are similarly encrypted.

5. The computer-implemented method of claim 3, in which producing the requested medical information that characterizes the group includes determining if the comparison provides at least a minimum number of matches.

6. The computer-implemented method of claim 1, in which producing the requested medical information that characterizes the group includes requesting a predefined portion of the medical information.

7. The computer-implemented method of claim 1, in which producing the requested medical information that characterizes the group includes determining if the group includes at least a minimum number of individuals.

8. The computer-implemented method of claim 1, in which the de-identified medical data is obtained from pharmacies.

9. The computer-implemented method of claim 1, wherein the produced medical information to report the characteristics of the group of individuals is provided absent authorization at any time from the individuals in the group of individuals.

10. The computer-implemented method of claim 1, wherein the medical information provided absent authorization from the group of individuals is in conformity with the Health Insurance Portability and Accountability Act (HIPAA).

11. A system comprising:

a hardware encryption server for producing a token for each individual included in a group composed of multiple individuals identified in a request received from a user device for medical information that characterizes the group absent authorization from the group of individuals wherein the identity of the individuals in the group is unattainable from the tokens; and

a data server for producing the requested medical information that characterizes the group from a set of de-identified medical data representing pre or post-encounter characteristics of the individuals by identifying tokens in the set of de-identified medical data and the associated de-identified medical data that correspond to the produced tokens, wherein an identity of each of the indi-

13

14

viduals is unattainable from the produced medical information, and the data server is configured to provide the produced medical information to report characteristics of the group of individuals absent authorization from individuals in the group of individuals.

12. The system of claim **11**, in which the produced token for each individual represents encrypted data that identifies the corresponding individual.

13. The system of claim **11**, in which the data server is configured to compare the produced token for each individual in the group to the tokens associated with the set of medical data.

14. The system of claim **13**, in which the tokens associated with the set of medical information and the produced token for each individual in the group are similarly encrypted.

15. The system of claim **13**, in which the data server is configured to determine if the comparison provides at least a minimum number of matches.

16. The system of claim **11**, in which the request represents a predefined portion of information to use for producing the medical information that characterizes the group.

17. The system of claim **11**, in which the encryption server is configured to determine if the group includes at least a minimum number of individuals.

18. The system of claim **11**, in which the de-identified medical data is obtained from pharmacies.

19. One or more computer readable storage devices storing instructions that are executable by a processing device, and upon such execution cause the processing device to perform operations comprising:

receiving a request from a user device for medical information that characterizes a group of individuals composed of multiple individuals from a set of de-identified medical data representing pre- or post-encounter characteristics of the individuals;

producing, by a hardware processor, tokens, wherein the identity of the individuals is unattainable from the tokens, each token being associated with an individual in the group, each token associable with a corresponding token in the set of de-identified medical data, and each token in the set of de-identified medical data being associated with de-identified medical data for a corresponding individual;

producing the requested medical information that characterizes the group of individuals by identifying the tokens in the set of de-identified medical data and the associated de-identified medical data that correspond to the produced tokens, wherein the produced medical information is absent personally identifiable information and has been produced without the individuals in the group having authorized production of such medical information that characterizes the group; and

providing the produced medical information to report the characteristics of the group of individuals absent authorization from the individuals in the group of individuals.

20. The one or more computer readable storage devices of claim **19**, in which producing the request token for each individual includes encrypting respective data that identifies each individual.

21. The one or more computer readable storage devices of claim **19**, in which producing the requested medical information that characterizes the group includes comparing the request token for each individual in the group to the tokens associated with the set of medical information data.

22. The one or more computer readable storage devices of claim **21**, in which the tokens associated with the set of medical information and the produced token for each individual in the group are similarly encrypted.

23. The one or more computer readable storage devices of claim **21**, in which producing medical information that characterizes the group includes determining if the comparison provides at least a minimum number of matches.

24. The one or more computer readable storage devices of claim **19**, in which producing medical information that characterizes the group includes requesting a predefined portion of the information.

25. The one or more computer readable storage devices of claim **19**, in which producing medical information that characterizes the group includes determining if the group includes at least a minimum number of individuals.

26. The computer readable storage devices of claim **19**, in which the de-identified medical data is obtained from pharmacies.

* * * * *

CERTIFICATE OF CORRECTION

PATENT NO.	: 9,323,892 B1	Page 1 of 1
APPLICATION NO.	: 14/082433	
DATED	: April 26, 2016	
INVENTOR(S)	: Andrew L. Paris, III	

It is certified that error appears in the above-identified patent and that said Letters Patent is hereby corrected as shown below:

Title Page, Column 2, item (57) (Abstract), Line 3, delete "pre" and insert -- pre- --, therefor.

Title Page, Column 2, item (57) (Abstract), Line 8, delete "pre" and insert -- pre- --, therefor.

In the Claims

Column 12, Line 63, In Claim 11, delete "pre" and insert -- pre- --, therefor.

Column 14, Line 21, In Claim 21, after "medical" delete "informational".

Signed and Sealed this
Ninth Day of August, 2016

Michelle K. Lee
Director of the United States Patent and Trademark Office